Representing Hybrid Compensatory Non-Compensatory Choice Set Formation in Semi-Compensatory Models

Sigal Kaplan
Technical University of Denmark
Department of Transport
Bygningstorvet 116 Vest
2800 Kgs. Lyngby - Denmark
Telephone: +45.45256559
Fax: +45.45936533
E-mail: siga@transport.dtu.dk

Shlomo Bekhor
Faculty of Civil and Environmental Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel
Tel: +972.4.829.2460
Fax: +972.4.829.5708
Email: sbekhor@technion.ac.il

Yoram Shiftan
Faculty of Civil and Environmental Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel
Tel: +972.4.829.2381
Fax: +972.4.829.5708
Email: shiftan@technion.ac.il

* Corresponding author

Revised version submitted for presentation to the 91th Annual Meeting of the Transportation Research Board, January 22-26, 2012, Washington D.C. and publication in the Transportation Research Record: Journal of the Transportation Research Board.

Submission date: November 12, 2011

Word count: 6,497 (text) + 4 × 250 (figures and tables) = 7,497

ABSTRACT

Semi-compensatory models represent a choice process consisting of an elimination-based choice set formation upon satisfying criteria thresholds and a utility-based choice. Current semi-compensatory models assume a purely non-compensatory choice set formation and hence do not support multinomial criteria that involve trade-offs among attributes at the choice set formation stage. This study proposes a novel behavioral paradigm comprising a hybrid compensatory non-compensatory choice set formation process, followed by compensatory choice. The behavioral paradigm is represented by a mathematical model that combines multinomial-response and ordered-response thresholds with a utility-based choice. The proposed model is applied to a stated preference experiment of off-campus rental apartment choices by students. Results demonstrate the applicability and feasibility of incorporating multinomial-response thresholds into semi-compensatory models.

1. INTRODUCTION

In complex choice situations, individuals engage in a decision process consisting of a sequence of choice set formation, in which alternatives are acceptable for consideration upon satisfaction of criteria thresholds, and utility-based choice among the considered alternatives (e.g., *1*). Traditionally, criteria are thought of as one-dimensional continuous or naturally ordered entities reflecting environmental and individual constraints that govern rational choice (*2*). Classic examples are physical and economic constraints (e.g., speed, size, time). However, some criteria are multidimensional entities that reflect both constraints and individual preferences, and their selection involves trade-offs among different attributes. For example, elimination of dwelling units on the basis of their location may be subject to evaluating the trade-off between accessibility and location amenities.

Decisions in transport and urban planning are typically represented by compensatory models accommodating attribute trade-offs. Residential choice is a classic example of a decision that is treated as a nested compensatory choice. For example, three-tier nested logit models describe the selection of tenure type, zone and dwelling type (*3*), and the selection of city, neighborhood and dwelling unit (*4*). Compensatory models are flexible in their ability to represent linear and non-linear utility functions, similarity across alternatives and taste heterogeneity, but present noteworthy limitations. Firstly, the assumption of utility-maximization as sole cognitive mechanism is behaviorally justified only in choice situations involving a few alternatives (*5*). Secondly, the assumption of individuals as being fully informed regarding all alternatives is unrealistic in choices among many alternatives (*6*). Last, neglecting information related to the choice set formation might result in mis-specified choice sets and lead to biased parameter estimates, lack of robustness in parameter estimates, and violations of the independence from irrelevant alternatives assumption (*6*, *7*).

Semi-compensatory models better comply with human decision making in choice situations entailing many alternatives, since they assume multiple cognitive mechanisms and partial information regarding the universal realm of alternatives (*8*). These models include a probabilistic two-stage choice process, consisting of an elimination-based choice set formation upon satisfying criteria thresholds, followed by a utility-maximization based choice. A disadvantage of current semi-compensatory models is the assumption of purely non-compensatory choice set formation followed by purely compensatory choice. Specifically, choice set formation is assumed to be governed by attribute-based elimination heuristics, and hence does not accommodate multinomial criteria.

Most existing semi-compensatory models consider only naturally ordered criteria for choice set formation. Borgers et al. (*9*) and Gensch (*10*) use ordinal performance criteria. Ben Akiva and Boccara (*1*) and Bierlaire et al. (*11*) employ availability constraints. Morikawa (*12*) applies information availability and minimum attraction constraints. Başar and Bhat (*7*) utilize an overall utility threshold. Swait (*13*), Cantillo and Ortúzar (*14*) and Cantillo et al. (*15*) consider time and cost criteria. Zheng and Guo (*16*) implement a distance-based criterion. Castro et al. (*17*) consider time-based criteria.

Attempts to represent multinomial criteria in addition to naturally ordered ones are conducted only by Swait (*18*) and Kaplan et al. (*8*). Swait (*18*) considers car-rental agencies and gasoline return policy as criteria in addition to price and size, and Kaplan et al. (*8*) specify apartment sharing and neighborhood as criteria in addition to price. However, both studies transform the multinomial criteria to naturally ordered ones prior to model estimation. Swait (*18*) transforms the multinomial criteria, namely the gasoline return policy criterion consisting of the three alternatives "prepay full tank", "return level" and "fill premium", into three binary criteria, namely "not prepay gas", "not return gas level rented", and "not pay gas premium". Kaplan et al. (*8*) transform the multinomial criteria into ordered criteria by considering the performance of the alternatives with respect to a single aspect. For example, in the apartment sharing criterion vacant apartments are assumed to be better than shared apartments of the same features. Both approaches lead to information loss since they fail to consider trade-offs among various aspects of the same criterion.

This study proposes the development and estimation of a semi-compensatory model that incorporates multinomial-response and ordered-response criteria. The model development includes the proposition of a novel behavioral paradigm that involves a hybrid compensatory non-compensatory (HCNC) choice set formation process, followed by compensatory choice. The behavioral paradigm is represented by a mathematical model that combines multinomial-response and ordered-response thresholds with a utility-based choice. Hence, the model represents a choice set formation driven by individual constraints and preferences.

The proposed semi-compensatory model is applied to a stated preference experiment of off-campus rental apartment choices by students, as an example of a decision involving both multinomial-response and ordered-response criteria.

The remainder of the paper is organized as follows. Section 2 presents the proposed behavioral paradigm and the corresponding mathematical model. Section 3 describes the data sample for model estimation and section 4 presents the estimation results. Last, section 5 draws conclusions and recommends further research.

2. MODEL FRAMEWORK

2.1 Behavioral paradigm

This study proposes a behavioral paradigm consisting of a HCNC choice set formation followed by compensatory choice. The proposed paradigm, illustrated in Figure 1, is an extension of the model suggested by Payne (*5*) and its theoretical foundation lies in the concept of "ecological rationality" suggested by Todd and Gigerenzer (*19*). According to this concept, the human mind stores a collection of human cognitive decision-making mechanisms, called "adaptive toolbox", from which decision strategies are retrieved in order to cope with the complexity of the choice environment while balancing decision accuracy and speed. The notion of "adaptive toolbox" suggests the possibility of combining various decision mechanisms in a flexible manner rather than adhering to a single mechanism of elimination followed by utility-maximization as suggested by Payne (5). The proposed paradigm is further supported by the conceptual model of information search and consideration set formation in a web-based environment proposed by Punj and Moore (*20*). This conceptual model (*20*) indicates that both additive strategies and elimination strategies are plausible for information search, and that electronic decision aids that are designed to help on-line consumers include applications for both information integration and information filtering. This notion can be easily understood by looking at the structure of commercial on-line shopping websites that implies the co-existence of compensatory and non-compensatory evaluations at the choice set formation stage. For example, on-line real-estate portals offer the possibility to search apartments by selecting both naturally ordered criteria thresholds, such as price and number of rooms, and multinomial criteria thresholds, such as city and apartment type.

The HCNC choice set formation includes three consecutive steps. The first step is the application of the simplification-by-aggregation heuristic (*21*) in order to characterize the universal realm of alternatives while balancing accuracy and cognitive effort. The simplification-by-aggregation heuristic utilizes the property of near-decomposability to reduce the dimensionality of complex systems. Specifically, when individuals identify a hierarchical structure or a similarity pattern across basic units of a complex system with respect to a certain dimension, they can replace the details of the basic units with a small number of aggregate units. This procedure, which is analogous to the statistical procedure of factor analysis, allows individuals to reduce the system complexity to a manageable level, since often individuals are unable to perform the computations when the system is represented in terms of its basic components. The simplification-by-aggregation heuristic is useful when gathering full and highly accurate information about the universal realm is costly, time consuming and cognitively effortful, since it allows characterizing the complex system at a simplified aggregate level on the basis of partial information. When the complex system is a large universal realm of alternatives, individuals simplify the choice situation by aggregating alternatives into groups that share similar traits. At each aggregation level, the alternatives are discrete and are characterized by a vector of attributes, although the number of attributes decreases as the level of aggregation increases. For example, in residential choice individuals may simplify the universal realm of dwellings by referring to the spatial dimension and aggregating the alternatives by neighborhoods. While dwelling units are characterized by their structural features and location amenities, only location amenities are retained at the neighborhood level. Hence, only information regarding location amenities is gathered by the individual in the case that the spatial dimension serves to simplify the universal realm of alternatives.

The second step is the evaluation of the aggregate groups in a compensatory process in order to account for trade-offs among attributes of aggregate units. The cognitive effort involved in evaluating aggregate units is lower than the cognitive effort involved in evaluating basic units, since the number of both aggregate units and considered attributes decrease as the aggregation level increases. The evaluation of aggregate units reflects the preference structure of individuals based on their knowledge and perceptions regarding the simplified universal realm of alternatives.

The third step is the application of the conjunctive heuristic (*22*) while combining two criteria types. The first type comprises naturally ordered criteria that derive from intrinsic individual constraints (either self-inflicted or externally-imposed). The second type consists of multinomial criteria that derive from individual preferences regarding the simplified universal realm of alternatives. While criteria deriving from intrinsic constraints are related solely to individual characteristics, criteria deriving from individual preferences are linked to the perception of the simplified universal realm formed in the mind of each individual at the previous steps. Specifically, the individual treats aggregation dimensions as multinomial elimination criteria and the aggregate units along each dimension as thresholds. Hence, upon the compensatory evaluation of aggregate units at the second step, the aggregate units that do not meet a tolerable utility threshold are discarded from further consideration.

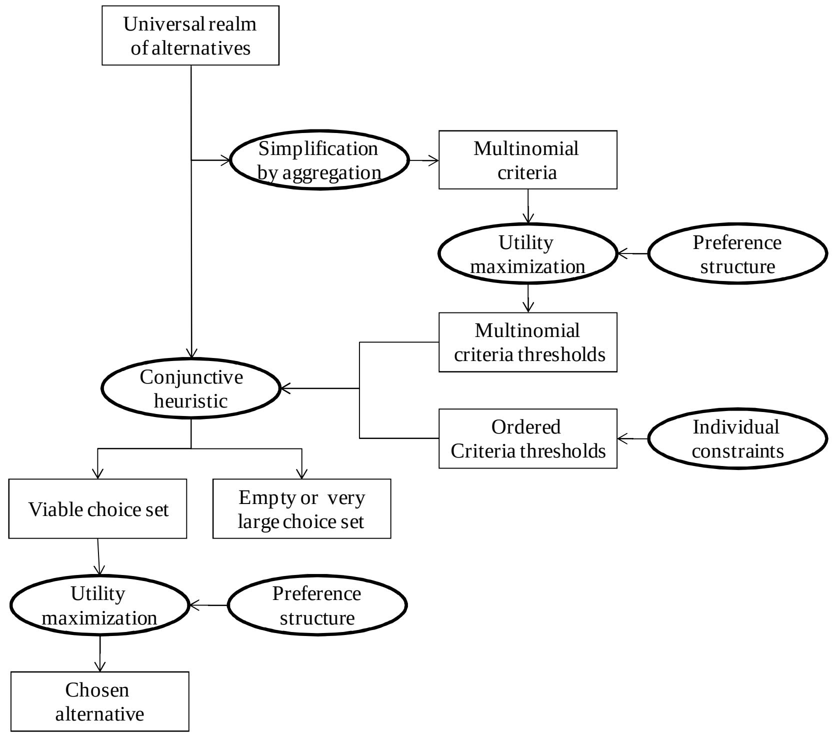

FIGURE 1 HCNC choice set formation followed by compensatory choice.

2.2 Mathematical formulation

The probability $P_q(i|G)$ of individual q $(q=1, 2,\ldots,Q)$ to choose alternative i, as Q individuals face the same universal realm G, is expressed as (*8*):

$$P_q(i|G) = P_q(i|S)P_q(S|G) \tag{1}$$

where $P_q(S|G)$ is the probability that individual q delimits the universal realm G to a viable choice set S, and $P_q(i|S)$ is the probability that individual q chooses alternative i from S. The equation is derived from Manski's formulation (*23*), but differs from the original expression by relying on observed choice sets in addition to choice outcomes. This allows reducing the number of possible choice sets to those actually chosen, hence avoiding the summation over all the theoretically possible choice sets in Manski's (*23*) original formula.

The proposed model assumes a HCNC choice set formation followed by utility-maximization. The HCNC choice set formation is based on the assumption that individuals apply the simplification-by-aggregation heuristic in order to represent the universal realm of alternatives in a crude though efficient manner. For parsimony reasons, the current mathematical formulation is based on two assumptions. The first assumption is that the population is homogeneous with respect to the simplified universal realm following the application of the simplification-by-aggregation heuristic. Namely, individuals group the alternatives in the universal realm in the same manner. The second assumption is that, while the simplification-by-aggregation heuristic is latent, its outcomes are observable. Hence, researchers possess full information regarding the simplified universal realm as perceived by the population.

The outcomes of the simplification-by-aggregation heuristic are the aggregate groups of alternatives, and individuals can utilize their preference structure regarding the simplified universal realm in order to delimit it to a viable choice set. Specifically, individuals retain in their viable choice sets only alternatives included within their most preferred aggregate group. The utility-based selection of the aggregate group g of alternatives from N aggregate groups depends on a vector Y_g of attributes of the aggregate group g, a vector Z_{qg} of characteristics of individual q associated with group g, and interaction terms Y_gZ_{qg} between attributes of the aggregate group and individual characteristics. The utility function is assumed to be linear in parameters:

$$U_{qg} = \delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg} + \eta_{qg} \qquad (2)$$

where δ, λ and γ are vectors of parameters to be estimated, and η_{qg} is the vector of error terms. The probability P_{qg} to choose group g from N groups of alternatives within the universal realm G is represented by a multinomial logit (MNL) model:

$$P_{qg} = F\left(\delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg}\right) = \frac{\exp\left(\delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg}\right)}{\sum_{n=1}^{N} \exp\left(\delta' Y_n + \lambda' Z_{qn} + \gamma' Y_n Z_{qn}\right)} \qquad (3)$$

In addition to narrowing down the universal realm of alternatives according to individual preferences regarding the perceived simplified universal realm, individuals can further delimit the considered set of alternatives by applying elimination criteria that represent their intrinsic constraints. Individual constraints are naturally ordered variables that depend solely on individual characteristics. For these criteria, the threshold t^*_{kq} is expressed as a function of a vector of individual characteristics Z_{kq}, a vector of coefficients to be estimated α_k, and an error term ε_{kq}:

$$t^*_{kq} = \alpha'_k Z_{kq} + \varepsilon_{kq} \qquad (4)$$

Assuming that the error term ε_{kq} for criterion k is identically and independently distributed (i.i.d.) normal across individuals, the ordered probit model represents the probability that individual q selects the threshold t^*_{kq}:

$$P_q\left(\theta_{(m-1)_k} < t^*_{kq} \le \theta_{m_k}\right) = \Phi\left(\theta_{(m-1)_k} - \alpha'_k Z_{kq}\right) - \Phi\left(\theta_{m_k} - \alpha'_k Z_{kq}\right) \qquad (5)$$

The probability to select a choice set S derives from the selection probability of a combination of the criteria corresponding to individual constraints and aggregate groups of alternatives that represent the preference of the individual regarding the perceived simplified universal realm. In the case that the error terms η_{qg} and ε_{kq} are uncorrelated, the probability to select the choice set S is simply the multiplication of the probabilities to select an aggregate group of alternatives and a threshold of an individual constraint. However, in the case that the error terms η_{qg} and ε_{kq} are correlated, the multiplication of the separate probabilities is inappropriate. The joint selection of an aggregate group of alternatives and a naturally ordered threshold to delimit the universal realm to a viable choice set can be represented by the joint multinomial ordered-response probit model developed by Bhat (*24*).

Following Bhat (*24*), the joint selection of an aggregate group and a threshold follows a bivariate normal distribution. In order to construct the bivariate normal distribution, the multinomial choice among N aggregate groups of alternatives is transformed into N binary choices. Let R_{qg} be a dummy variable that equals one if the aggregate group g is chosen by individual q and zero otherwise. Define:

$$\nu_{qg} = \left\{ \max_{n=1,\dots,N;n\neq g} U_{qn} \right\} - \eta_{qg} \tag{6}$$

U^*_{qg} is the utility of choosing the aggregate group g by individual q:

$$U^*_{qg} = \delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg} - \nu_{qg} \tag{7}$$

R_{qg} equals one if U^*_{qg} is positive and zero otherwise. The error term v_{qg} is transformed into a standard normal random variable v^*_{qg}:

$$\nu^*_{qg} = \Phi^{-1}\left[F_g\left(\nu_{qg} \right) \right] \tag{8}$$

where Φ is the standard normal distribution function and F is the multinomial logit distribution function of v_{gq}. Assuming that both v_{qg} and ε_{kq} are independently distributed across individuals, and considering that both v^*_{qg} and ε_{kq} follow a normal distribution, a bivariate distribution can be specified:

$$\Phi_2\left[\nu^*_{qg}, \varepsilon_{kq}, \rho_q \right] = \Phi_2\left[\Phi^{-1}\left[F\left(\nu_{qg} \right) \right], \varepsilon_{kq}, \rho_q \right] \tag{9}$$

where Φ_2 represents the bivariate normal distribution and ρ_g is the correlation between the aggregate group g and the individual constraint. The joint probability of choosing group g and threshold t^*_{kq} by individual q is expressed as follows:

$$\begin{aligned} P_q\left(g, \theta_{(m-1)_k} < t^*_{kq} \le \theta_{m_k} \right) &= \Phi_2\left(\Phi^{-1}\left[F\left(\delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg} \right) \right], \theta_{(m-1)_k} - \alpha'_k Z_{kq}, \rho_g \right) \\ &- \Phi_2\left(\Phi^{-1}\left[F\left(\delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg} \right) \right], \theta_{m_k} - \alpha'_k Z_{kq}, \rho_g \right) \end{aligned} \tag{10}$$

The corresponding likelihood is:

$$\begin{aligned} L_q\left(S \mid G \right) = \prod_{n=1}^{N} \Bigg\{ \prod_{m_k=1}^{M_k} \Bigg[& \Phi_2\left(\Phi^{-1}\left[F\left(\delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg} \right) \right], \theta_{(m-1)_k} - \alpha'_k Z_{kq}, \rho_g \right) \\ & - \Phi_2\left(\Phi^{-1}\left[F\left(\delta' Y_g + \lambda' Z_{qg} + \gamma' Y_g Z_{qg} \right) \right], \theta_{m_k} - \alpha'_k Z_{kq}, \rho_g \right) \Bigg]^{d_{qm_k}} \Bigg\}^{d_{qg}} \end{aligned} \tag{11}$$

where d_{qmk} is an indicator function that is equal to one if individual q selects the threshold category m of criterion k and zero otherwise, and d_{qg} is an indicator function that is equal to one if individual q selects the aggregate group g and zero otherwise.

The utility-based choice stage is represented by considering a linear utility function with an i.i.d. Gumbel error term. The utility-based selection of alternative i from the viable choice set S depends on a vector X_i of attribute values of alternative i, a vector Z_{qi} of individual characteristics associated with alternative i, and interaction terms $X_g Z_{qi}$ between attributes of alternative i and individual characteristics:

$$U_{qg} = \beta' X_i + \omega' Z_{qi} + \mu' X_g Z_{qi} + \zeta_{qi} \tag{12}$$

where β, ω and μ are vectors of parameters to be estimated, and ζ_{qi} is the vector of error terms. The probability for individual q to choose alternative i from J alternatives within the viable choice set S is represented by an MNL model:

$$P_q\left(i \mid S\right) = \frac{\exp\left(\beta' X_i + \omega' Z_{qi} + \mu' X_g Z_{qi}\right)}{\sum_{j\in S}^{J} \exp\left(\beta' X_j + \omega' Z_{qj} + \mu' X_g Z_{qj}\right)} \tag{13}$$

The corresponding likelihood is:

$$L_q\left(i \mid S\right) = \prod_{j\in S}\left[\frac{\exp\left(\beta' X_i + \omega' Z_{qi} + \mu' X_g Z_{qi}\right)}{\sum_{j\in S}^{J} \exp\left(\beta' X_i + \omega' Z_{qi} + \mu' X_g Z_{qi}\right)}\right]^{d_{qi}} \tag{14}$$

where d_{qi} equals unity if threshold alternative i is chosen by individual q and zero otherwise.

The error terms of the HCNC choice set formation and the compensatory choice are assumed to be independent. Accordingly, the combined unconditional log-likelihood for a population of Q individuals who choose their most preferred alternative i from their viable choice set S_q can be written as:

$$LL = \sum_{q=1}^{Q} \ln\left[L_q\left(i \mid S_q\right) L_q\left(S_q \mid G\right)\right] \tag{15}$$

The coefficients of both the HCNC choice set formation and the utility-maximization are estimated simultaneously by using maximum likelihood estimation with a stepwise procedure. In the first step, log-likelihood is maximized with respect to the threshold selection parameters assuming an independent error structure. In the second step, the model parameters are held fixed and the log-likelihood function is maximized with respect to the correlation parameters. In the third step, the parameters of the threshold selection and the correlation parameters are held fixed and the log-likelihood function is maximized with respect to the utility parameters. Finally, the parameters from the third step are used as starting values for the full-information maximum likelihood estimation of the log-likelihood function.

3. DATA

The data sample for model estimation consists of 858 students in the city of Haifa, Israel, who participated in a web-based experiment of rental apartment choice. Kaplan et al. (*25*) discuss the survey design, detailing its theoretical foundation, the synthetic dataset construction and the questionnaire design.

Participants delimited the universal realm of alternatives to a viable choice set from which they chose their preferred alternative. Specifically, participants searched a synthetically generated apartment dataset by a list of pre-defined criteria threshold values and ranked their three most preferred apartments from the resulting choice set. The apartment dataset was constructed on the basis of a statistical analysis of local real-estate databases and consisted of rental apartments characterized by 18 attributes. The search criteria were apartment sharing, neighborhood and monthly rent price.

Notably, during the experiment respondents could revise their tolerated thresholds until they obtained a satisfactory choice set from which they chose their preferred apartment. Moreover, respondents that did not find their ideal apartment either chose a sub-optimal alternative or aborted the survey. Unfortunately, the revision of thresholds and the possibility to abort the survey were not captured in the current survey and are left for further research.

A supplemental questionnaire collected participants' socio-economic characteristics, attitudes and perceptions about issues relevant to rental apartment choice. Participants were also asked to evaluate perceived location amenities of neighborhoods, since public records of quantitative information about neighborhood amenities in Haifa are unavailable. Specifically, questions investigated job accessibility for students, leisure accessibility, public open space availability, campus accessibility by public transport, and perceived car travel time to campus. All the items were expressed on a seven-point Likert scale ranging from 1 (very poor) to 7 (excellent), except the perceived car travel time that was expressed in minutes.

The web-server automatically recorded the participants' answers to the questionnaire and typing actions during the two-stage choice experiment. The three most preferred apartments for each of the 858 students served for model estimation. Hence, the data sample consists of a pool of 2,574 observations of choice outcomes and their corresponding thresholds. Among respondents, the average age is 26.6 year (standard deviation = 3.8), 60.7% are male, 29.1% are married, 44.5% have an available car on a daily basis, 73.1% live in the city of Haifa, 33.1% live in the dorms, 12.0% live with their parents, 18.4% rent an apartment alone or with a roommate and 36.4% live with their spouse. When requested about monthly expenses as a proxy for income, most participants (57.0%) declared 750 USD or less, 12.9% answered up to 1,000 USD, 26.3% indicated higher values, and 3.7% refused to answer. A comparison with available public records shows that the sample is representative of the student population in terms of gender distribution, median age, and residential arrangements (*25*).

4. RESULTS

Table 1 presents the estimation results for the model combining an HCNC choice set formation and compensatory choice with an independent error structure at the utility-maximization stage (HCNC-MNL). The model is presented while considering alternatively independent and correlated thresholds.

Three criteria are represented in the estimated model: apartment sharing, neighborhood and monthly rent price. These criteria were ranked as the most important rental apartment attributes in a preliminary survey and were utilized for searching the database by the entire population sample.

The model structure is based on the assumption that individuals simplified the universal realm by aggregating the alternatives into eight mutually exclusive groups on the basis of apartment sharing and neighborhood. Apartment sharing differentiates between vacant and shared apartments and four neighborhoods (i.e., Carmel, Neveshanan, Remez, Nesher) are considered. Figure 2 illustrates the considered aggregate groups of alternatives whose selection is represented by a multinomial logit model.

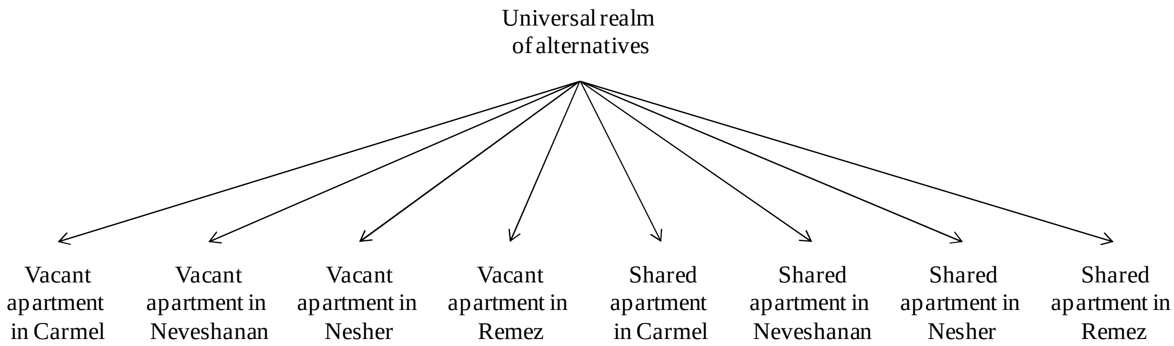

FIGURE 2 Aggregate groups simplifying the universal realm of alternatives.

The price threshold is best described by an ordered probit model with 11 categories (200, 250,…,700). The choice of an aggregate group and a price threshold leads to the formation of 83 non-empty choice sets.

The relevant universal realm for the population sample contains all the 400 apartments available in the generated synthetic database for the four aforementioned neighborhoods. A priori availability constraints are not imposed on the alternatives but, for each ranked choice, the alternatives with higher priority ranking are excluded from the choice set as respondents could not rank the same alternative twice. Notably, repeated choices of the same individual are treated as independent and data are treated as pooled cross-sectional data.

The variable specification for the choice set formation stage is based on the assumption that individual constraints are related to individual characteristics and perceptions, and that the choice among aggregate groups of alternatives is related to their perceived utility. Hence, the perceived utility is related to both the attributes of the aggregate groups of alternatives and individual characteristics.

The selection among the aggregate groups of alternatives is related to students' job accessibility, leisure activity accessibility, public open space availability, campus accessibility by public transport, and perceived car travel time to campus. These attributes are entered as generic variables in the model (i.e., part-worth utilities are constant across all alternatives). In addition, alternative specific constants are associated with Carmel, Nesher and Remez with respect to Neveshanan. An interaction term between studying in the Faculty of Medicine and residing in Carmel is estimated, since the Carmel neighborhood provides higher accessibility to the medical campus. Interaction terms between car availability and Carmel, Nesher or Remez are estimated, since Carmel is located far from the main campus and Nesher and Remez are located on a hilly terrain, while Neveshanan is adjacent to campus and on a level terrain. Finally, individual characteristics are associated with the four aggregate groups entailing vacant apartments: socio-economic variables (i.e., marital status, gender, age, monthly expenses), transport related variables (i.e., car availability), study-place preferences, and current residential location and arrangements.

The selection of the price threshold depends on socio-economic characteristics (i.e., age, marital status, monthly expenses, income source), transport related variables (i.e., car availability, trip frequency to the campus), apartment search experience, current residential location and living arrangement. In addition, latent constructs designed to measure price knowledge and travel preferences are considered. Observed personal characteristics are directly used as explanatory variables, whereas perceptions and preferences are incorporated in the model after performing factor analysis.

The model specification assumes that the utility-maximization choice depends on the attributes of the apartments included within the choice set. According to Kaplan et al. (*7*), the inclusion of interaction terms between individual characteristics and apartment attributes for this particular case study results in only a minimal increase of the model goodness-of-fit. Hence, parsimony considerations suggest the current model specification not to include interaction terms between individual characteristics and apartment attributes. Considered apartment attributes include price, structural features (i.e., number of rooms, renovation status, floor, security bars), location amenities (i.e., view, parking, noise level, proximity to campus), electrical appliances (i.e., air conditioning system and solar water heater), number of roommates and smoking policy.

The first three parts of table 1 describe the determinants of the HCNC choice set formation. The fourth part presents the relative importance of apartment attributes at the choice stage, given the viable choice set. The interpretation of the model results is provided in the following sub-sections.

TABLE 1 HCNC-MNL Model Estimation Results

Variable	Description	HCNC-MNL Independent thresholds		HCNC-MNL Correlated thresholds	
		est.	t-stat.	est.	t-stat.
	Apartment type threshold coefficients				
Marital status	Single [a]	-	-	-	-
	Married	2.058	9.39	2.049	9.86
Gender	Female [a]	-	-	-	-
	Male	-0.847	-6.02	-0.839	-6.39
Age	Years	0.024	2.45	0.024	2.54
Car availability	Monthly/weekly [a]	-	-	-	-
	Daily	0.692	4.63	0.691	4.94
Trip frequency to campus	Monthly/weekly [a]	-	-	-	-
	Daily	-0.601	-4.37	-0.606	-4.53
On-campus studying to benefit from teacher-student communication factor		-0.142	-3.61	-0.142	-3.57
Monthly expenses	< $750 [a]	-	-	-	-
	$ 750 - 1000	0.504	2.36	0.511	2.41
	$1000 - 1750	0.777	3.74	0.776	3.87
Current residential arrangement	Dormitories [a]	-	-	-	-
	Parents	-	-	-	-
	Roommates	-1.189	-5.91	-1.156	-6.08
	Alone	1.352	5.14	1.351	5.09
	Spouse	1.730	9.63	1.729	10.47
Current residential location	Haifa city [a]	-	-	-	-
	Haifa suburbs	-0.504	-1.87	-0.504	-1.87
	Haifa outskirts	-0.639	-5.01	-0.631	-2.85

	Center of Israel	-	-	-	-
Neighborhood specific constant	Neve Shaanan [a]	-	-	-	-
	Carmel	-2.239	-13.78	-2.243	-12.25
	Nesher	-1.585	-13.31	-1.577	-13.20
	Remez	-1.401	-14.03	-1.407	-14.82
Accessibility to job opportunities		0.144	5.92	0.141	5.57
Availability of leisure opportunities		0.106	4.04	0.107	4.02
Availability of gardens and open spaces		0.324	14.35	0.325	13.62
Arrival ease to campus by public transport		0.008	0.40	0.007	0.37
Travel time to campus by car		-0.035	-4.58	-0.036	-4.32
Interaction between daily car availability and residence in Carmel		1.378	9.81	1.389	8.89
Interaction between daily car availability and residence Nesher		0.718	5.44	0.725	5.33
Interaction between daily car availability and residence Remez		0.296	2.25	0.294	2.13
Interaction between residence in Carmel and studying in the medical campus		1.348	6.01	1.334	5.81
Interaction between residence Nesher and studying in faculties adjacent to the campus's eastern gate		0.516	4.08	0.521	3.93
Price threshold coefficients					
Marital status	Single [a]	-	-	-	-
	Married	0.588	7.41	0.571	5.83
Gender	Female [a]	-	-	-	-
	Male	-0.201	-3.71	-0.196	-3.48
Age	Years	0.035	4.76	0.034	3.55
Monthly expenses	< $ 500 [a]	-	-	-	-
	$ 500-750	0.366	6.18	0.369	5.07
	$ 750-1500	0.444	5.81	0.457	5.21
	> $1500	0.585	5.74	0.593	4.78
Income source	None/scholarship [a]	-	-	-	-
	Part-time job	0.091	1.82	0.091	1.50
	Full-time job	-	-	-	-
Car availability	Monthly/weekly [a]	-	-	-	-
	Daily	0.243	4.06	0.236	3.66
Price-knowledge factor		0.083	5.64	0.083	4.80
Apartment search experience	#3 apartment changes [a]	-	-	-	-
	> 4 apartment changes	-0.251	-2.34	-0.252	-2.36
Trip frequency to campus	Monthly/weekly [a]	-	-	-	-
	Daily	-0.347	-5.79	-0.352	-5.27
Current residential arrangement	Dormitories [a]	-	-	-	-
	Roommates	-0.399	-5.11	-0.406	-4.96
	Alone/parents	0.244	3.51	0.244	3.04
	Spouse	0.567	7.30	0.558	6.72
Current residential location	Haifa –low/medium	-	-	-	-

	class neighborhoods [a]				
	Haifa – upper class neighbourhoods	0.251	3.62	0.249	2.77
	Haifa Suburbs	-	-	-	-
	Haifa outskirts	-	-	-	-
	Center of Israel	0.217	3.05	0.216	2.30
Non-motorized modes preference factor		-0.033	-2.34	-0.033	-1.82
Travel minimization preference factor		-0.041	-2.69	-0.041	-2.13
Cut-off points	200 [a]	-	-	-	-
	250	-0.185	-0.80	-0.184	-0.62
	350	0.277	1.20	0.282	0.95
	350	0.549	2.37	0.548	1.84
	400	0.754	3.26	0.758	2.53
	450	1.209	5.20	1.237	4.10
	500	1.639	7.04	1.677	5.55
	550	2.221	9.56	2.280	7.52
	600	2.448	10.53	2.541	8.33
	650	2.753	11.84	2.993	9.64
	700	2.883	12.39	3.304	10.26
Correlations across thresholds					
Vacant apartment in Carmel and price		-	-	-0.207	-2.39
Vacant apartment in Neveshanan and price		-	-	-0.322	-3.71
Vacant apartment in Nesher and price		-	-	-0.060	-0.71
Vacant apartment in Remez and price		-	-	-0.088	-1.16
Shared apartment in Carmel and price		-	-	0.166	1.07
Shared apartment in Neveshanan and price		-	-	0.485	9.97
Shared apartment in Nesher and price		-	-	0.326	5.39
Shared apartment in Remez and price		-	-	0.384	6.68
Choice stage coefficients given the selected choice set					
Rent price (monthly)		-0.004	-8.46	-0.004	-8.58
Number of rooms		0.686	16.30	0.683	16.21
Number of roommates		-0.570	-8.10	-0.567	-7.96
Walking time to campus		-0.083	-21.19	-0.084	-21.32
Quiet apartment		1.358	29.28	1.341	28.97
Parking		0.408	7.05	0.414	7.13
Floor		-0.116	-5.46	-0.114	-5.34
Smoking allowed		-0.258	-3.87	-0.255	-3.77
Security bars		0.187	4.12	0.184	4.02
Stunning view		0.243	4.88	0.242	4.83
Renovated apartment		0.467	9.59	0.470	9.60
Air conditioner		0.673	14.03	0.666	13.70
Solar water heater		0.573	9.51	0.573	9.33
Number of observations			2574		2574
Number of parameters			67		75

Log-likelihood at zero	-29268.825	-29268.825
Log-likelihood at estimates	-16578.954	-15465.313
McFadden's adjusted R^2	0.431	0.469

Note: [a] base category

4.1 HCNC choice set formation

Determinants of the choice among aggregate groups of alternatives related to apartment sharing and neighborhood.

The propensity of selecting an aggregate group of alternatives increases according to several neighborhood characteristics: (i) greater accessibility to job opportunities; (ii) greater availability of leisure activities; (iii) greater availability of open spaces; (iv) greater ease of arrival to campus by car; (v) lower car travel time from the neighborhood to campus.

Results suggest that students have an inherent preference for the Neveshanan neighborhood and also a higher inherent preference for Remez, Nesher and Neveshanan relatively to the Carmel neighborhood. Possible reasons may be the existence of higher student concentrations in these neighborhoods, the ease of arrival to campus by non-motorized modes and the lower municipal taxes.

The interaction terms between car availability and propensity to delimit the universal realm to apartments in Carmel, Nesher and Remez have positive parameter estimates. The parameter values increase with the distance from campus and the change from level to hilly terrain. Specifically, Neveshanan is adjacent to campus and on a level terrain, Remez and Nesher are adjacent to campus but on a hilly terrain, and Carmel is both located far from campus and characterized by a hilly terrain. Hence, students with daily car availability have higher propensity to select less accessible neighborhoods than other students, and this tendency strengthens with increasing difficulty to reach the campus by non-motorized modes.

A positive parameter estimate is obtained for the interaction term between studying in faculties adjacent to the campus eastern gate and residing in Nesher. Possibly, due to the hilly terrain near the campus, students who study in faculties located in the eastern side of the campus prefer to reside in the only neighborhood in proximity of the eastern gate.

Individual characteristics are associated exclusively with the four groups of vacant apartments. The propensity of selecting an aggregate group including only vacant apartments increases according to the progression of the respondents' lifecycle in terms of age, marital status and monthly expenses, and to travel independence related to daily car availability.

The propensity to delimit the universal realm of alternatives to an aggregate group including only shared apartments increases with daily trips to campus and the preference to study there in order to benefit from teacher-student interaction. Possibly, respondents who spend more time on campus than in their apartment prefer to ease apartment chores by sharing them.

The propensity to retain a certain apartment type increases with the reference to the status-quo, as respondents currently residing with roommates have a greater propensity to retain an aggregate group of alternatives including only shared apartments, while respondents residing alone or with their spouse tend to retain an aggregate group of alternatives consisting only of vacant apartments.

4.2 Threshold selection determinants for monthly rent price

The propensity to select higher price thresholds increases according to: (i) progression of the students' lifecycle and socio-economic status, as respondents who are married or have higher monthly expenses tend to select higher price thresholds; (ii) self-reported price knowledge; (iii) reference to the status-quo, as respondents who currently rent an apartment alone or with their spouse have a greater propensity to select higher price thresholds than respondents who reside either in the dormitories or with roommates.

The propensity to select higher price thresholds decreases according to: (i) habit to travel daily to campus, likely related to shorter time spent in the apartment with respect to the time spent on campus; (ii) greater apartment search experience, likely reflecting a greater propensity to undergo the burden of replacing the status-quo apartment with a more cost-efficient one.

The likelihood ratio test value (LR = 2227.3), which compares the model with an independent error structure at the choice set formation stage versus the model with correlations between the choice of an aggregate group of alternatives and the price threshold, exceeds the critical chi-square value for eight correlation parameters at the 0.001 significance level. Thus, the null assumption of independent error structure at the choice set formation stage can be rejected. As expected, the correlation parameters between the four aggregate groups of vacant apartments and the price threshold are negative, indicating a positive correlation between delimiting the universal realm to vacant apartments and selecting a higher price threshold. Similarly, the correlation parameters between the four aggregate groups of shared apartments and the price threshold are positive, indicating a negative correlation between delimiting the universal realm to shared apartments and selecting a higher price threshold. Correlations between the price threshold and delimiting the universal realm to vacant apartments in the neighborhoods of Nesher and Remez are not statistically significant. The correlation between the price threshold and selecting a shared apartment in Carmel is also not statistically significant, although the reason may be the relatively low choice frequency (36 respondents) of this aggregate group of alternatives.

4.3 Utility-maximization choice

For each respondent, the choice of an aggregate group of alternatives according to apartment sharing and neighborhood determines whether the chosen apartment is vacant or shared and the neighborhood in which it is located. Hence, apartment sharing and neighborhood do not vary within the viable choice set. Monthly rent price and number of roommates vary within the viable choice set and hence serve as explanatory variables at the utility-maximization stage.

The propensity of renting an apartment increases according to an increase in terms of: (i) quality of structural features, such as greater number of rooms and renovation; (ii) availability of security bars; (iii) availability of cost-efficient electrical appliances such as solar water heater; (iv) availability of an air conditioning system, a necessity in the hot Mediterranean climate; (v) parking availability; (vi) environmental location amenities, such as quietness and a nice view.

The propensity of renting an apartment decreases according to an increase in terms of: (i) apartment monthly rent; (ii) floor number, as apartments located on lower floors are preferred, possibly due to the scarcity of elevators in the two neighborhoods; (iii) number of roommates; (iv) distance from campus, as an increase in walking time to the campus decreases the attractiveness of the apartment; (v) roommates' pro-smoking policy, possibly since 87.0% of the respondents in the sample are non-smokers.

4.4 Comparison with the traditional approach

Table 2 presents the estimation results for the MNL model for the same data in order to compare the proposed approach with the traditional residential choice modeling approach. The main difference between the proposed HCNC-MNL model and the MNL model is in the treatment of the choice set. In the HCNC-MNL model, alternatives that do not meet the criteria thresholds are eliminated from further consideration at the choice stage. In the MNL model, the full universal realm of 400 alternatives is considered at the choice stage. The two models are compared in terms of their in-sample predictive ability.

TABLE 2 MNL model Estimation Results

Variable	Description	MNL	
		est.	t-stat.
Rent price (monthly)		-0.008	-22.83
Number of rooms		0.770	20.32
Number of roommates		-0.692	-18.15
Walking time to campus		-0.009	-15.38
Quiet apartment		1.369	31.57
Parking		0.047	0.92
Floor		-0.235	-11.96
Smoking allowed		-0.121	-2.09
Security bars		0.259	5.98
Stunning view		0.128	2.63
Renovated apartment		0.501	11.03
Air conditioner		0.259	6.07
Solar water heater		0.244	4.47
Number of observations			2574
Number of parameters			13
Log-likelihood at zero			-15422.018
Log-likelihood at estimates			-14023.487
McFadden's adjusted R^2			0.090

The predictive ability is typically assessed by the percent correctly predicted (PCP), namely the percent of observations in which the actual choice outcomes corresponds to the alternative with the highest probability. This approach is criticized for its inability to incorporate uncertainty and to account for the magnitude of the relative difference in the choice probabilities (*26*). Hence, in addition to the traditional approach, the PCP is hereby assessed as the percent of observations in which the actual choice outcomes corresponds to one of the three most probable alternatives. This adjustment allows incorporating uncertainty and considering alternatives with very similar choice probabilities. Considering the traditional approach, the PCP of the MNL is 2.7%, while the PCP of the HCNC-MNL is 9.0%. The adjusted PCP for three alternatives is 11.3% for the MNL, versus 19.8% for the HCNC-MNL.

5. CONCLUSIONS AND FURTHER RESEARCH

This study proposes the development and estimation of a semi-compensatory model that incorporates both multinomial-response and ordered-response criteria. To the authors' knowledge, this is the first semi-compensatory model that relaxes the assumption of a purely non-compensatory choice set formation and allows trade-offs to occur at both the choice set formation stage and the choice stage.

The empirical analysis uses a data set from a stated preference experiment of off-campus rental apartment choices by students. Results indicate that: (i) the hypothesized model has an excellent goodness-of-fit, thus supporting the proposed behavioral paradigm; (ii) the compensatory evaluation of the aggregate apartment groups at the choice set formation stage is related to neighborhood characteristics and individual socio-economic characteristics; (iii) accounting for the correlations between the multinomial and the ordered-thresholds greatly improves the model goodness-of-fit; (iv) the utility-based choice from the viable choice set is related to apartment structural features, number of roommates, presence of electrical appliances, and location amenities. Notably, while the results derive from a particular survey interface, the survey design and the apartment database utilized were based on actual data drawn from real-estate databases (*8, 25*) and hence the analysis is sufficiently general to provide information regarding local student residential choice in the study area.

The proposed approach can be applied to various spatial and transportation related choices entailing many alternatives, such as trip destination, long distance travel, recreational destination choice, workplace choice, car rental choice and route choice. Nevertheless, a necessary condition for the model application consists of data collection about choice set formation. Specifically, the data needs for model estimation include threshold elicitation, either directly or indirectly by requesting information about the considered non-chosen alternatives. While this information is not collected in traditional surveys, it may be retrieved from on-line transaction in the real-estate, car-rental, tourism and airline industries. Moreover, with the advancement of semi-compensatory models it would be beneficial to reconsider the inclusion of choice set formation information in traditional surveys.

Several possible directions exist for further development of the proposed model. Firstly, the current study assumes i.i.d. error terms across groups at the choice set formation stage and across alternatives in the choice stage. The i.i.d. assumption implies restrictive substitution patterns across alternatives and taste homogeneity across the population (*26*). Thus, a potential research direction is the relaxation of the i.i.d. assumption by using the mixed logit model for representing the compensatory choice due to its ability to represent random taste heterogeneity and flexible substitution patterns across alternatives. Secondly, the current model formulation assumes population homogeneity with respect to the simplified universal realm and researchers' full information about the structure of the simplified universal realm. Possibly, these two assumptions could be relaxed by allowing the simplified structure of the universal realm to vary probabilistically across the population. Last, the current model formulation employs Bhat' (*24*) multinomial ordered-response probit model. An alternative approach for representing the HCNC choice set formation could be to estimate copula-based models, due to their ability to represent a wide range of multidimensional distributions involving discrete and continuous variables.

ACKNOWLEDGMENTS

The authors are grateful to four anonymous reviewers, who provided insightful comments to improve this paper.

REFERENCES

1. Ben-Akiva, M.E., and B. Boccara. Discrete choice models with latent choice sets. *International Journal of Research in Marketing, Vol. 12*, 1995, pp. 9-24.
2. Simon, H.A. A behavioral model of rational choice. *The Quarterly Journal of Economics, Vol. 69*, 1955, pp. 99-118.
3. Yates, J., and D.F. Mackay. Discrete choice modeling of urban housing markets: a critical review and an application. *Urban Studies, Vol. 43*, 2006, pp. 559-581.
4. Chattopadhyay, S. The effectiveness of McFadden's nested logit model in valuing amenity improvement. *Regional Science and Urban Economics, Vol. 30*, 2000, pp. 23-43.
5. Payne, J.W. Task complexity and contingent processing in decision making: an information search and protocol analysis. *Organisational Behavior and Human Performance, Vol. 16*, 1976, pp. 366-387.
6. Shocker, A.D., M.E. Ben-Akiva, B. Boccara, and P. Nedungadi. Consideration set influences on consumer decision-making and choice: issues, models and suggestions. *Marketing Letters, Vol. 2*, 1991, pp. 187-197.
7. Başar, G., and C. R. Bhat. A parameterized consideration set model for airport choice: an application to the San Francisco Bay Area. *Transportation Research Part B: Methodological, Vol. 38*, 2004, pp. 889-904.
8. Kaplan, S., S. Bekhor, and Y. Shiftan. A two-stage model jointly revealing the determinants of non-compensatory conjunctive choice set formation and compensatory choice. *Transportation Research Record: Journal of the Transportation Research Board, Vol. 2134*, 2009, pp.153-163.
9. Borgers, A., H.Timmermans, and J. Veldhuisen. A hybrid compensatory-non-compensatory model for residential preference structures. *Journal of Housing and the Built Environment, Vol. 1*, 1986, pp. 227-234.
10. Gensch, D.H. A two-stage disaggregate attribute choice model. *Marketing Science, Vol. 6*, 1987, pp. 223-239.
11. Bierlaire, M., R. Hurtubia, and G. Flötteröd. An analysis of the implicit choice set generation using the Constrained Multinomial Logit model. *Proceedings of the 89th TRB Annual Meeting*, 2010, Washington D.C.
12. Morikawa, T. A hybrid probabilistic choice set model with compensatory and non-compensatory choice rules. *Proceedings of the 7th WCTR*, 1995, Sydney, Australia.
13. Swait, J.A. Choice set generation within the generalized extreme value of discrete choice models. *Transportation Research Part B: Methodological, Vol. 35*, 2001, pp. 643-666.

14. Cantillo, V., and J.d.D. Ortúzar. A semi-compensatory choice model with explicit attribute thresholds of perception. *Transportation Research Part B: Methodological, Vol. 39*, 2005, pp. 641-657.

15. Cantillo, V., B. Heydecker, and J.d.D. Ortúzar. A discrete choice model incorporating thresholds for perception in attribute values. *Transportation Research Part B: Methodological, Vol. 40,* 2006, pp. 807-825.

16. Zheng, J. and J.Y. Guo. A destination choice model incorporating choice set formation. *Proceedings of the 87th TRB Annual Meeting*, 2008, Washington, D.C.

17. Castro, M., F. Martinez, and M. Munizaga. Calibration of a logit discrete choice model with elimination of alternatives. *Proceedings of the International Choice Modelling Conference,* 2009, Harrogate, U.K.

18. Swait, J.A. non-compensatory choice model incorporating attribute cutoffs. *Transportation Research Part B: Methodological, Vol. 35*, 2001, pp. 903-928.

19. Todd, P.M., and G. Gigerenzer. Bounding rationality to the world. *Journal of Economic Psychology, Vol. 24*, 2003, pp. 143-165.

20. Punj, G., and R. Moore. Information search and consideration set formation in a web-based store environment. *Journal of Business Research Vol. 62*, 2009, pp. 644-650.

21. Simon, H.A. Can there be a science of complex systems? In Y. Bar-Yam (Ed.), *Unifying Themes in Complex Systems. Proceedings from the International Conference on Complex Systems*, 1999, Cambridge, MA: Perseus Press.

22. Einhorn, H.J. The use of nonlinear, non-compensatory models in decision making. *Psychological Bulletin, Vol. 73*, 1970, pp. 221-230.

23. Manski, C. The structure of random utility models. *Theory and Decision*, *Vol. 8*, 1977, pp. 229-254.

24. Bhat, C.R. Work travel mode choice and number of Non-work commute stops. *Transportation Research Part B: Methodological, Vol. 31*, 1997, pp. 41-54.

25. Kaplan, S., S. Bekhor, and Y. Shiftan. Web-based survey design for unraveling semi-compensatory choice in transport and urban planning. *Transportation planning and technology (forthcoming),* 2012.

26. Train, K. Discrete Choice Methods with Simulation. Cambridge University Press, 2009, New York, USA.